WHAT IS CLAIMED IS:

1. A method for processing search results generated based on a query, the method comprising:

a) accepting the search results;

b) accepting information derived from the query;

c) generating a set of final search results from the accepted search results based on the accepted information.

2. The method of claim 1 wherein the act of generating a set of final search results includes

i) determining whether or not a candidate search result is similar to a search result already in the set of final search results; and

ii) if it is determined that the candidate search result is similar to a search result already in the set of final search results, then not adding the candidate search result to the set of final search results.

3. The method of claim 2 wherein the act of determining whether or not a candidate search result is similar to a search result already in the set of final search results includes

A) extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information;

B) extracting at least a part of the search result already in the set of final search results that is relevant to the information derived from the query, thereby generating second query-relevant information; and

C) determining whether or not the first query-relevant information is similar to the second query-relevant information,

wherein, if the first query-relevant information is determined to be similar to the second query-relevant information, then determining the candidate search result to be similar to the search result already in the set of final search results, and

wherein, if the first query-relevant information is determined not to be similar to the second query-relevant information, then determining the candidate search result not to be similar to the search result already in the set of final search results.

4. The method of claim 3 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) defining a window as a first predetermined number of characters;

2) applying the window to various parts of a document corresponding to the candidate search result;

3) for each of the various parts of the document to which a window is applied,

- determining the number of keywords in the current part of the document to determine a hit count;

4) ranking the various parts of the document to which a window is applied based on its associated hit count; and

5) taking a second predetermined number of the highest ranking various parts of the document to define at least a part of the first query-relevant information.

5. The method of claim 3 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting a document corresponding to the candidate search result to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at least one of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least one of the query keywords, then

adding the segment to the first query-relevant information.

6. The method of claim 5 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into sentences.

7. The method of claim 5 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into paragraphs.

8. The method of claim 3 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting a document corresponding to the candidate search result to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at least a predetermined number of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least the predetermined number of the query keywords, then adding the segment to the first query-relevant information.

9. The method of claim 8 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into sentences.

10. The method of claim 8 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into paragraphs.

11. The method of claim 3 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting a document corresponding to the search result to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at a predetermined number of different ones of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least the predetermined number of different ones of the query keywords, then adding the segment to the first query-relevant information.

12. The method of claim 11 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into sentences.

13. The method of claim 11 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into paragraphs.

14. The method of claim 1 wherein the act of generating a set of final search results includes

i) determining whether or not a candidate search result is similar to a search result already in the set of final search results; and

ii) adding the search results to the set of final search results only if it is determined that the candidate search result is not similar to any search results already in the set of final search result.

15. The method of claim 14 wherein the act of determining whether or not a candidate search result is similar to a search result already in the set of final search results includes

A) extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information;

B) extracting at least a part of the search result already in the set of final search

results that is relevant to the information derived from the query, thereby generating second query-relevant information; and

C) determining whether or not the first query-relevant information is similar to the second query-relevant information,

wherein, if the first query-relevant information is determined to be similar to the second query-relevant information, then determining the candidate search result to be similar to the search result already in the set of final search results, and

wherein, if the first query-relevant information is determined not to be similar to the second query-relevant information, then determining the candidate search result not to be similar to the search result already in the set of final search results.

16. The method of claim 15 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) defining a window as a first predetermined number of characters;

2) applying the window to various parts of a document corresponding to the candidate search result;

3) for each of the various parts of the document to which a window is applied,

- determining the number of keywords in the current part of the document to determine a hit count;

4) ranking the various parts of the document to which a window is applied based on its associated hit count; and

5) taking a second predetermined number of the highest ranking various parts of the document to define at least a part of the first query-relevant information.

17. The method of claim 15 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting a document corresponding to the search result to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at least one of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least one of the query keywords, then adding the segment to the first query-relevant information.

18. The method of claim 17 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into sentences.

19. The method of claim 17 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into paragraphs.

20. The method of claim 15 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting a document corresponding to the candidate search result to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at a predetermined number of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least the predetermined number of the query keywords, then adding the segment to the first query-relevant information.

21. The method of claim 20 wherein the act of segmenting a document corresponding to the candidate search result to

define a plurality of segments, segments the document into sentences.

22. The method of claim 20 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into paragraphs.

23. The method of claim 15 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the candidate search result that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting a document corresponding to the candidate search result to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at a predetermined number of different ones of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least the predetermined number of different ones of the query keywords, then adding the segment to the first query-relevant information.

24. The method of claim 23 wherein the act of segmenting a document corresponding to the candidate search result to

define a plurality of segments, segments the document into sentences.

25. The method of claim 23 wherein the act of segmenting a document corresponding to the candidate search result to define a plurality of segments, segments the document into paragraphs.

26. A method for determining whether or not a first document corresponding to a first search result is similar to a second document corresponding to a second search result, the method comprising:

a) accepting a query that generated the first and second search results;

b) extracting at least a part of the first document that is relevant to the information derived from the query, thereby generating first query-relevant information;

c) extracting at least a part of the second document that is relevant to the information derived from the query, thereby generating second query-relevant information; and

d) determining whether or not the first query-relevant information is similar to the second query-relevant information,

wherein, if the first query-relevant information is determined to be similar to the second query-relevant information, then determining the first document to be similar to the second document, and

wherein, if the first query-relevant information is determined not to be similar to the second

query-relevant information, then determining the first document not to be similar to the second document.

27. The method of claim 26 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the first document that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) defining a window as a first predetermined number of characters;

2) applying the window to various parts of the first document;

3) for each of the various parts of the first document to which a window is applied,

- determining the number of keywords in the current part of the first document to determine a hit count;

4) ranking the various parts of the first document to which a window is applied based on its associated hit count; and

5) taking a second predetermined number of the highest ranking various parts of the first document to define at least a part of the first query-relevant information.

28. The method of claim 26 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the first documents that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting the first document to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at least one of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least one of the query keywords, then adding the segment to the first query-relevant information.

29. The method of claim 28 wherein the act of segmenting the first document to define a plurality of segments, segments the first document into sentences.

30. The method of claim 28 wherein the act of segmenting the first document to define a plurality of segments, segments the first document into paragraphs.

31. The method of claim 26 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the first document that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting the first document to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at a predetermined number of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least the predetermined number of the query keywords, then adding the segment to the first query-relevant information.

32. The method of claim 31 wherein the act of segmenting the first document to define a plurality of segments, segments the first document into sentences.

33. The method of claim 31 wherein the act of segmenting the first document to define a plurality of segments, segments the first document into paragraphs.

34. The method of claim 26 wherein the information derived from the query includes query keywords, and

wherein the act of extracting at least a part of the first document that is relevant to the information derived from the query, thereby generating first query-relevant information includes:

1) segmenting the first document to define a plurality of segments;

2) for each of the segments, determining whether or not the segment includes at a predetermined number of different ones of the query keywords; and

3) for each of the segments, if it was determined that the segment includes at least the predetermined number of different ones of the query keywords, then adding the segment to the first query-relevant information.

35. The method of claim 34 wherein the act of segmenting the first document to define a plurality of segments, segments the document into sentences.

36. The method of claim 34 wherein the act of segmenting the first document to define a plurality of segments, segments the document into paragraphs.

37. A machine-readable medium including machine executable instructions which, when executed by a machine, processes search results generated based on a query by:

a) accepting the search results;

b) accepting information derived from the query;

c) generating a set of final search results from the accepted search results based on the accepted information.

38. The machine-readable medium of claim 37 including further machine executable instructions which, when executed by a machine, generate the set of final search results by

i) determining whether or not a candidate search result is similar to a search result already in the set of final search results; and

ii) adding the candidate search result to the set of final search results only if it is determined that the candidate search result is not similar to any search results already in the set of final search result.

39. The machine-readable medium of claim 38 including further machine executable instructions which, when executed by a machine determine whether or not a search result is similar to a search result already in the set of final search results by

A) extracting at least a part of the search result that is relevant to the information derived from the query, thereby generating first query-relevant information;

B) extracting at least a part of the search result already in the set of final search results that is relevant to the information derived from the query, thereby generating second query-relevant information; and

C) determining whether or not the first query-relevant information is similar to the second query-relevant information,

wherein, if the first query-relevant information is determined to be similar to the second query-relevant information, then determining the search results to be similar to the search result already in the set of final search results, and

wherein, if the first query-relevant information is determined not to be similar to the second query-relevant information, then determining the search

results not to be similar to the search result already in the set of final search results.

40. An apparatus for processing search results generated based on a query, the apparatus comprising:

a) a storage facility for storing the search results and for storing information derived from the query; and

b) a final results generator for generating a set of final search results from the search results stored in the storage facility based on the information stored in the storage facility.

41. The apparatus of claim 40 wherein the final set generator includes

i) a similarity determination facility for determining whether or not a search result is similar to a search result already in the set of final search results; and

ii) means for adding the search results to the set of final search results only if the similarity determination facility determines that the search result is not similar to any search results already in the set of final search result.

42. The method of claim 41 wherein the similarity determination facility includes

A) a query-relevant extraction facility for extracting at least a part of the search result that is relevant to the information derived from the query, thereby generating

first query-relevant information, and for extracting at least a part of the search result already in the set of final search results that is relevant to the information derived from the query, thereby generating second query-relevant information; and

B) a query-relevant similarity determination facility for determining whether or not the first query-relevant information is similar to the second query-relevant information,

wherein, if the query-relevant similarity determination facility determines that the first query-relevant information is similar to the second query-relevant information, then the similarity determination facility determines the search result to be similar to the search result already in the set of final search results, and

wherein, if the query-relevant similarity determination facility determines that the first query-relevant information is not similar to the second query-relevant information, then the similarity determination facility determines the search result not to be similar to the search result already in the set of final search results.

43. A storage facility including at least one machine-readable medium storing information comprising:

a) ranked query results;

b) query-relevant parts of documents corresponding to the ranked query results; and

c) a final set of query results,

wherein the final set of query results is a sub-set of the ranked query results, and

wherein the final set of query results does not include any two query results corresponding to documents that have similar query-relevant parts.

44. The storage facility of claim 43, the stored information further comprising:

d) documents corresponding to the ranked query results.

45. A method for processing search results generated based on a query, the method comprising:

a) accepting the search results;

b) accepting information derived from the query;

c) accepting documents associated with the search results;

d) extracting portions of the documents associated with the search results based on the information derived from the query to generate query-relevant information for each of the documents; and

e) generating a set of final search results from the accepted search results based on the query-relevant information.